

```

1 ATGCTGGGGG CAGTGAAGG CCCCAGGTGG AAGCAGGCGG AGGACATTAG
51 AGACATCTAC GACTTCCGAG ATGTTCTGGG CACGATCAAG CACCCCAACA
101 TTGTAGCCCT GGATGACATC TATGAGAGTG GGGGCCACCT CTACCTCATC
151 ATGCAGCTGG TGTCGGGTGG GGAGCTCTTT GACCGTATTG TGGAAAAAGG
201 CTTCTACACG GAGCGGGACG CCAGCCGCCT CATCTTCCAG GTGCTGGATG
251 CTGTGAAATA CCTGCATGAC CTGGGCATTG TACACCGGGA TCTCAAGCCA
301 GAGAATCTGC TGTACTACAG CCTGGATGAA GACTCCAAAA TCATGATCTC
351 CGACTTTGGC CTCTCCAAGA TGGAGGACCC GGGCAGTGTG CTCTCCACCG
401 CCTGTGGAAC TCCGGGATAC GTGGCCCCTG AAGTCCTGGC CCAGAAGCCC
451 TACAGCAAGG CTGTGGATTG CTGGTCCATA GGTGTCATCG CCTACATCTT
501 GCTCTGCGGT TACCCTCCCT TCTATGACGA GAATGATGCC AAACCTTTTG
551 AACAGATTTT GAAGGCCGAG TACGAGTTTG ACTCTCCTTA CTGGGACGAC
601 ATCTCTGACT CTGCCAAGA TTTTCATCCG CACTTGATGG AGAAGGACCC
651 AGAGAAAAGA TTCACCTGTG AGCAGGCCTT GCAGCACCCA TGGATTGCAG
701 GAGATACAGC TCTAGATAAG AATATCCACC AGTCGGTGAG TGAGCAGATC
751 AAGAAGAACT TTGCCAAGAG CAAGTGAAG CAAGCCTTCA ATGCCACGGC
801 TGTGGTGCGG CACATGAGGA AACTGCAGCT GGGCACCAGC CAGGAGGGGGC
851 AGGGGCAGAC GGCGAGCCAT GGGGAGCTGC TGACACCAGT GGCTGGGGGGC
901 CCGGCAGCTG GCTGTTGCTG TCGAGACTGC TCGTGGAGC CGGGCACAGA
951 ACTGTCCCCC AACTGCCCC ACCAGCTCTA G (SEQ ID NO:1)

```

#### FEATURES:

Start Codon: 1  
Stop Codon: 979

#### Homologous proteins:

##### Top 10 BLAST Hits

|  | Score | E     |
|--|-------|-------|
| CRA 18000004983962 /altid=gi 4502553 /def=ref NP_003647.1  (NM...  | 661   | 0.0   |
| CRA 18000004936440 /altid=gi 3122310 /def=sp Q63450 KCC1_RAT CA... | 642   | 0.0   |
| CRA 223000002652742 /altid=gi 15928726 /def=gb AAH14825.1 AAH14... | 641   | 0.0   |
| CRA 18000005144641 /altid=gi 3114436 /def=pdb 1A06  calmodulin...  | 556   | e-157 |
| CRA 18000004932361 /altid=gi 406113 /def=gb AAA19670.1  (L24907... | 548   | e-155 |
| CRA 117000066864297 /altid=gi 9966875 /def=ref NP_065130.1  (NM... | 470   | e-131 |
| CRA 149000126089143 /altid=gi 14422219 /def=emb CAC41379.1  (AL... | 398   | e-109 |
| CRA 224000007378166 /altid=gi 16755792 /def=gb AAL28100.1 AF428... | 398   | e-109 |
| CRA 114000110934306 /altid=gi 14196445 /def=ref NP_065172.1  (N... | 398   | e-109 |
| CRA 18000005191499 /altid=gi 4007153 /def=emb CAA19296.1  (AL02... | 398   | e-109 |

FIGURE 1A

**Blast hits to dbEST:**

| CRA Number          | gi Number   | Score           | Expect |
|---------------------|-------------|-----------------|--------|
| CRA 58000099505996  | gi 12943070 | 1459 bits (736) | 0.0    |
| CRA 164000139365918 | gi 12675371 | 1415 bits (714) | 0.0    |
| CRA 58000099322782  | gi 12899184 | 1215 bits (613) | 0.0    |
| CRA 78000169264025  | gi 14067900 | 1134 bits (572) | 0.0    |
| CRA 225000015220001 | gi 18523306 | 1130 bits (570) | 0.0    |
| CRA 225000015219990 | gi 18523305 | 1130 bits (570) | 0.0    |
| CRA 225000001633290 | gi 15750044 | 1102 bits (556) | 0.0    |
| CRA 61000077034868  | gi 14446412 | 1063 bits (536) | 0.0    |
| CRA 335000490524629 | gi 8278341  | 995 bits (502)  | 0.0    |
| CRA 146000055060127 | gi 10205334 | 954 bits (481)  | 0.0    |
| CRA 61000077035673  | gi 14446457 | 890 bits (449)  | 0.0    |
| CRA 1000488750278   | gi 5128333  | 884 bits (446)  | 0.0    |
| CRA 225000001678100 | gi 15752845 | 882 bits (445)  | 0.0    |
| CRA 222000012165952 | gi 18781967 | 850 bits (429)  | 0.0    |
| CRA 224000004550264 | gi 15947485 | 718 bits (362)  | 0.0    |
| CRA 162000005790240 | gi 9185548  | 706 bits (356)  | 0.0    |
| CRA 225000000831163 | gi 15496148 | 317 bits (160)  | 6e-85  |

**EXPRESSION INFORMATION FOR MODULATORY USE (library source):**

| gi Number   | Organ | Tissue Type    |
|-------------|-------|----------------|
| gi 10205334 | eye   | retinoblastoma |
| gi 15496148 | brain | hypothalamus   |

2050E0-20006001

**FIGURE 1B**

1 MLGAVEGPRW KQAEIRDIT DFRDVLGTIK HPNIVALDDI YESGGHLYLI  
 51 MQLVSGGELF DRIVEKGFYT ERDASRLIFQ VLDVAVKYLHD LGIVHRDLKP  
 101 ENLLYYSLDE DSKIMISDFG LSKMEDPGSV LSTACGTPGY VAPEVLAQKP  
 151 YSKAVDCWSI GVIAYILLCG YPPFYDENDA KLFEQILKAE YEFDSPYWDD  
 201 ISDSAKDFIR HLMEKDPEKR FTCEQALQHP WIAGDTALDK NIHQSVSEI  
 251 KKNFAKSKWK QAFNATAVVR HMRKLQLGTS QEGQGQTASH GELLTPVAGG  
 301 PAAGCCCRDC CVEPGTELSP TLPHQL (SEQ ID NO:2)

# FEATURES:

## Functional domains and key regions:

### Prosite results:

PDOC00001 PS00001 ASN\_GLYCOSYLATION

N-glycosylation site

264-267 NATA

PDOC00004 PS00004 CAMP\_PHOSPHO\_SITE

CAMP- and cGMP-dependent protein kinase phosphorylation site

219-222 KRFT

PDOC00005 PS00005 PKC\_PHOSPHO\_SITE

Protein kinase C phosphorylation site

Number of matches: 3

|   |         |     |
|---|---------|-----|
| 1 | 28-30   | TIK |
| 2 | 70-72   | TER |
| 3 | 204-206 | SAK |

PDOC00006 PS00006 CK2\_PHOSPHO\_SITE

Casein kinase II phosphorylation site

Number of matches: 9

|   |         |      |
|---|---------|------|
| 1 | 55-58   | SGGE |
| 2 | 70-73   | TERD |
| 3 | 107-110 | SLDE |
| 4 | 122-125 | SKME |
| 5 | 204-207 | SAKD |
| 6 | 236-239 | TALD |
| 7 | 245-248 | SVSE |
| 8 | 279-282 | TSQE |
| 9 | 289-292 | SHGE |

PDOC00007 PS00007 TYR\_PHOSPHO\_SITE

Tyrosine kinase phosphorylation site

62-69 RIVEKGFY

PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 3

|   |         |        |
|---|---------|--------|
| 1 | 128-133 | GSVLST |
| 2 | 283-288 | GQGQTA |
| 3 | 299-304 | GGPAAG |

PDOC00100 PS00108 PROTEIN\_KINASE\_ST

Serine/Threonine protein kinases active-site signature

93-105 IVHRDLKPENLLY

FIGURE 2A

### Membrane spanning structure and domains:

| Helix | Begin | End | Score | Certainty |
|-------|-------|-----|-------|-----------|
| 1     | 127   | 147 | 0.883 | Putative  |
| 2     | 155   | 175 | 1.515 | Certain   |
| 3     | 288   | 308 | 0.649 | Putative  |

### BLAST Alignment to Top Hit:

>CRA|18000004983962 /altid=gi|4502553 /def=ref|NP\_003647.1|  
(NM\_003656) calcium/calmodulin-dependent protein kinase I  
[Homo sapiens] /org=Homo sapiens /taxon=9606 /div=PRI  
/dataset=nraa /length=370  
Length = 370

Score = 661 bits (1688), Expect = 0.0

Identities = 326/370 (88%), Positives = 326/370 (88%), Gaps = 44/370 (11%)

Frame = +3

Query: 126 MLGAVEGPRWKQAEDIRDIYDFRDVLGT----- 209  
MLGAVEGPRWKQAEDIRDIYDFRDVLGT  
Sbjct: 1 MLGAVEGPRWKQAEDIRDIYDFRDVLGTGAFSEVILAEDKRTQKLVAIKCIAKEALEGKE 60

Query: 210 -----IKHPNIVALDDIYESGGHLYLIMQLVSGGELFDRIVEKGFYTERDASR 353  
IKHPNIVALDDIYESGGHLYLIMQLVSGGELFDRIVEKGFYTERDASR  
Sbjct: 61 GSMENEIAVLHKIKHPNIVALDDIYESGGHLYLIMQLVSGGELFDRIVEKGFYTERDASR 120

Query: 354 LIFQVLDAVKYLHDLGIVHRDLKPENLLYSLDEDSKIMISDFGLSKMEDPGSVLSTACG 533  
LIFQVLDAVKYLHDLGIVHRDLKPENLLYSLDEDSKIMISDFGLSKMEDPGSVLSTACG  
Sbjct: 121 LIFQVLDAVKYLHDLGIVHRDLKPENLLYSLDEDSKIMISDFGLSKMEDPGSVLSTACG 180

Query: 534 TPGYVAPEVLAQKPYSKAVDCWSIGVIAYILLCGYPPFYDENDAKLFEQILKAEYFDSP 713  
TPGYVAPEVLAQKPYSKAVDCWSIGVIAYILLCGYPPFYDENDAKLFEQILKAEYFDSP  
Sbjct: 181 TPGYVAPEVLAQKPYSKAVDCWSIGVIAYILLCGYPPFYDENDAKLFEQILKAEYFDSP 240

Query: 714 YWDDISDSAKDFIRHLMKDPKRFCEQALQHPWIAAGDTALDKNIHQSVSEQIKKNFAK 893  
YWDDISDSAKDFIRHLMKDPKRFCEQALQHPWIAAGDTALDKNIHQSVSEQIKKNFAK  
Sbjct: 241 YWDDISDSAKDFIRHLMKDPKRFCEQALQHPWIAAGDTALDKNIHQSVSEQIKKNFAK 300

Query: 894 SKWKQAFNATAWVRHMRKLQLGTSQEGQGQTASHGELLTPVAGGPAAGCCCRDCCVEPGT 1073  
SKWKQAFNATAWVRHMRKLQLGTSQEGQGQTASHGELLTPVAGGPAAGCCCRDCCVEPGT  
Sbjct: 301 SKWKQAFNATAWVRHMRKLQLGTSQEGQGQTASHGELLTPVAGGPAAGCCCRDCCVEPGT 360

Query: 1074 ELSPTLPHQL 1103  
ELSPTLPHQL  
Sbjct: 361 ELSPTLPHQL 370 (SEQ ID NO:4)

FIGURE 2B

# Hmmer search results (Pfam):

Scores for sequence family classification (score includes all domains):

| Model   | Description                                | Score  | E-value | N |
|---------|--|--------|---------|---|
| PF00069 | Eukaryotic protein kinase domain           | 271.6  | 1.1e-77 | 1 |
| CE00022 | CE00022 MAGUK_subfamily_d                  | 119.7  | 3.1e-35 | 1 |
| CE00359 | E00359 bone_morphogenetic_protein_receptor | 6.5    | 0.36    | 1 |
| CE00031 | CE00031 VEGFR                              | 4.3    | 0.2     | 1 |
| PF01496 | V-type ATPase 116kDa subunit family        | 1.8    | 7.5     | 1 |
| CE00292 | CE00292 PTK_membrane_span                  | -89.8  | 0.0011  | 1 |
| CE00287 | CE00287 PTK_Eph_orphan_receptor            | -96.6  | 0.044   | 1 |
| CE00291 | CE00291 PTK_fgf_receptor                   | -123.4 | 0.11    | 1 |
| CE00286 | E00286 PTK_EGF_receptor                    | -151.4 | 0.095   | 1 |
| CE00290 | CE00290 PTK_Trk_family                     | -204.5 | 0.4     | 1 |
| CE00016 | CE00016 GSK_glycogen_synthase_kinase       | -271.7 | 0.12    | 1 |

Parsed for domains:

| Model   | Domain | seq-f | seq-t | hmm-f | hmm-t     | score     | E-value |
|---------|--------|-------|-------|-------|-----------|-----------|---------|
| PF01496 | 1/1    | 85    | 95    | ..    | 1 11      | [. 1.8    | 7.5     |
| CE00359 | 1/1    | 93    | 146   | ..    | 272 327   | .. 6.5    | 0.36    |
| CE00031 | 1/1    | 79    | 163   | ..    | 1053 1137 | .. 4.3    | 0.2     |
| CE00290 | 1/1    | 2     | 212   | ..    | 1 282     | [] -204.5 | 0.4     |
| CE00292 | 1/1    | 3     | 227   | ..    | 1 288     | [] -89.8  | 0.0011  |
| CE00291 | 1/1    | 1     | 230   | [.    | 1 285     | [] -123.4 | 0.11    |
| CE00287 | 1/1    | 2     | 230   | ..    | 1 260     | [] -96.6  | 0.044   |
| CE00286 | 1/1    | 1     | 230   | [.    | 1 263     | [] -151.4 | 0.095   |
| CE00022 | 1/1    | 30    | 232   | ..    | 75 283    | .. 119.7  | 3.1e-35 |
| PF00069 | 1/1    | 25    | 232   | ..    | 43 278    | .] 271.6  | 1.1e-77 |
| CE00016 | 1/1    | 1     | 302   | [.    | 1 433     | [] -271.7 | 0.12    |

FIGURE 2C

|      |            |             |             |             |             |
|------|------------|-------------|-------------|-------------|-------------|
| 1    | AAACCGACCT | TTGGCCTCTT  | GCCTGCCGTC  | CTAGTTGCAG  | GCTCTCTCCC  |
| 51   | CTAACCTGGA | CCCCAGCCAT  | CAAACCTCTG  | AGCCCCGCCA  | GTCACGTGAC  |
| 101  | ACCTCGGTCC | TTTTTTGGCCT | GTTTCCTTCA  | GGATCCCCGAT | TAACTTCTCT  |
| 151  | CCTCCCCAAT | TCCCTCTGCC  | CCCAATACCT  | CTAGGCACCA  | CCACCCGCTC  |
| 201  | TGAGGAGCAA | GTGTCTGGGG  | CTGAAGCCTC  | AGCTCCATCT  | TGCAGAGGAA  |
| 251  | CCGGGGCCTC | AGTCTTCCCA  | CCTGTCAAGT  | GGGGCCCACA  | CCCTGCGACC  |
| 301  | ACCTCCACTC | TCTTCATTGC  | CTAGTCTTGC  | CCGGTCCTTC  | CCCACTCCCT  |
| 351  | CACTCCCCCA | TCCCCCACCA  | GACTCCCGTG  | CAGTTCCAGG  | GCCTGTTTCC  |
| 401  | CTTCAGGGCA | CGGAGAAGGG  | AGACAGAGCC  | CTAAGGGAGG  | TCGCAGAACT  |
| 451  | GGTCTGAAAG | AAAATCCACC  | AGGCCACAGG  | GTGAGTTTGG  | CCGGCCTCTA  |
| 501  | GCTTCAGACA | GACGGGGTTC  | GAATCCTGCT  | TTGCTTCCGA  | CCACCCGCTG  |
| 551  | ATTTGAAAAT | CATCTCTCCG  | GGCCTCAGTT  | GTCCCCCTCTG | TGAAATGGAC  |
| 601  | CCCGCTTAAG | ACCAAGGGCG  | GGAAGCGTCC  | AGCAGGAGAT  | CTCTGACCAG  |
| 651  | AAGCAGGGAG | ATGGCCTCCA  | CCCGTGCCCC  | TTCCCCAGCC  | TTGGAGCGGT  |
| 701  | GCCTCGCCTC | CCAATCCCGG  | GTCCCTCCGC  | CGCAGGCTCC  | ACCTCCACTG  |
| 751  | ACATCAGAGC | CGCAGGCGGG  | CGGAGAGAGC  | CGCCGAGCCG  | AGCCGAGCCC  |
| 801  | CAGCTCCAGC | AAGAGCGCGG  | GCGGGTGGCC  | CAGGCACGCA  | GCGGTGAGGA  |
| 851  | CCGCGGCCAC | AGCTCGGCGC  | CAACCACCGC  | GGGCCTCCCA  | GCCAGCCCCG  |
| 901  | CGGCGGGGCA | GCCGCAGGTA  | CAGCCGGGCC  | CCCCATCCCT  | GCACCCCTGG  |
| 951  | GCGCTGCGTG | GGGGCGGTGG  | GAGCCCCTAG  | CCTCTGGGTA  | TCCTTTCCCA  |
| 1001 | AGGAGTGGCC | ACTGGGCACT  | CTCCCGGGCG  | GGCTGGACCC  | TGAGGGGCGAG |
| 1051 | GGCTGGGCCT | TTCTCCACCT  | CTGTCCCAGG  | CCCAGCAGGT  | GCCAGGCGGG  |
| 1101 | CCTATGGGAC | ACTGAGTGGG  | TAATAGAGAA  | GGGGGCCTGT  | GTGAGCGCCT  |
| 1151 | TCAGCTGGGC | CTGACTGGAA  | GGGCGTGGGC  | ATTTGGAGGT  | ATCCATGGGG  |
| 1201 | TGGGGGGGCT | TGCGGAGTGT  | ACTGTCTTAG  | GACAGGCGTG  | TGGGTGAGAC  |
| 1251 | ATGGGTGGAG | GATCTGGGAA  | TCTGTGTGTT  | TTTTGTTCCA  | GAGGGGTGTC  |
| 1301 | CACGTGTTTT | GTGTGCTGGT  | ATTTGGCTCT  | CAGGGTCTTA  | AGTCAGAGTT  |
| 1351 | AGGAGGGGGT | GTACAATTGT  | GTAAGTGGGA  | TGTTTGGAGT  | TAGGTGTGTA  |
| 1401 | AGGACTTGGG | GTTTGGTTTG  | GAATACAGGA  | GCTTCCAGGG  | GATGGGGTAG  |
| 1451 | AGGAGCTGGA | GGGTGTAGGG  | TACGTCTGGT  | ATATGAGGGT  | GTGTGTGTGT  |
| 1501 | GTGTCTGGGT | GTCATCTTGT  | GTGGGTGCGG  | GTGGATGTGT  | GTTTTGGGGT  |
| 1551 | GTAAGAGGGA | GCTGGGTGAG  | GGATGTTTGG  | ATGGACAGGC  | AGGTGTTCCG  |
| 1601 | GTGCAGGGCT | GTCTGGGGCA  | CTGTGTGGTG  | TGGACATGTG  | TGCTGATGTC  |
| 1651 | TGGGAGTACA | TGTATGATCA  | GGTGTACCGG  | GATGTGGATA  | CAAGGCGTAC  |
| 1701 | TGGATCTGGG | AGGCAGGTGT  | TTGAGTTCAG  | GGCTGTGGAG  | GGGGCTTGGT  |
| 1751 | GTGGCATGTC | TGCTACAGGG  | ATGTGTGTGG  | ATCTGTGAGG  | GTTGTATTTG  |
| 1801 | GTAGGCCTCC | ATGTGGGTTT  | CAGACTCTGC  | CTCTAGAGCT  | TACACTCGAG  |
| 1851 | TCTCCTTTCC | TAGAAGATTC  | TGCCCCCTGA  | TGGGTGGGCA  | GGGTCCCCTG  |
| 1901 | GGAAAAAGGT | CCTGTTCCAG  | GAGTGGAATC  | TCACACCAGA  | GGCCCTAGTC  |
| 1951 | AGGGCACCTT | CTCCTCATTG  | TCCCTTAGAG  | AAAAAGAGAG  | AAGGAAAGTG  |
| 2001 | CTCTCCCTGA | GGTCACAAAG  | CATGCTGGGC  | TCTGTTTTGG  | CCTCATCTGT  |
| 2051 | GGATGGGTTG | GGAGGCTGTG  | TTCTCTGAAT  | GGGGCCCATT  | CTGGCTTCAT  |
| 2101 | ATTGGAAGTA | CCAGCCAAGG  | CCATTTCGATG | GCCTTTGCCC  | TCAGCAAGCT  |
| 2151 | TAGCTGGGGG | CCCCAGGCCA  | GGTGTCAATTA | GGGCCTCTGG  | AGCCAGCCTC  |
| 2201 | TACCTAACTC | CAACCTCAGT  | CTCCCCATTC  | TTCATCTGAT  | AAATGGGAGA  |
| 2251 | GAACCTCCAC | CCTCTCCTGC  | TGGATGAGAC  | AGACCTCAGC  | AGAGGAAGGG  |
| 2301 | CCAGGCTGGA | TAGGGTTAGA  | TGGGGCCAGG  | AAGGGACAGA  | GTGAGCAGGA  |
| 2351 | CCATTTCTCA | TGCTCCCGGG  | ACCCAGATGG  | GGAGTCAGGA  | GGGAGAGGTC  |
| 2401 | TGGGGAGCTC | CAGCTGTGGC  | TGTTGTTGCT  | GTGGTAACAG  | TGCAGAAAGA  |
| 2451 | GCTATTTAAA | AATGTGGCTG  | AGATGTTGCT  | GGAAGCCCAG  | GCTGCTGGAA  |
| 2501 | ACCTGATTTT | GGAGAGGCCG  | GGGAGTCGGG  | GGAAGGAGGA  | GGGAAAGGAG  |
| 2551 | ACACCCCAGC | AATCCCCAGG  | GTGGGGCGGG  | GACATCACTG  | GTTCTGGGGA  |
| 2601 | CAGGGGGATC | CTCCAGGCTT  | CTACCAGCTG  | CTCTGGGGGT  | TTATCTGTTG  |
| 2651 | TACTGCCAGA | AGTCAGGGTT  | TCCCTAGGTG  | CTTGGAATTTG | GATAGGGGGA  |

FIGURE 3A

|      |            |            |            |             |             |
|------|------------|------------|------------|-------------|-------------|
| 2701 | AAACTGGGAA | GAGAACTAGA | ATAAATGAAT | GAATGAATGC  | ATGACTTTGT  |
| 2751 | TAAATAAAGA | ATTTTGCTGC | CACTGTGAAA | GGTTTTTCTC  | TAGGCATGAG  |
| 2801 | AATTTGCTGA | ATGTTGAATA | AACAAATGAA | TGTTTGTTGA  | ATGATTTTGT  |
| 2851 | CAAATGGATG | AATCAAGGAT | GAATAAATGC | AGGTTGAATG  | ACTGAATGGG  |
| 2901 | GCCTGCAGTA | AATTTCCAGA | CAGAGGGCTG | GGCTCTGCTG  | AGTCTCCTCC  |
| 2951 | TTCCATTCTC | CTTACAGGAG | CCCTGGCTGT | GGTCGGGGGG  | CAGTGGGCCA  |
| 3001 | TGCTGGGGGC | AGTGGAAGGC | CCCAGGTGGA | AGCAGGCGGA  | GGACATTAGA  |
| 3051 | GACATCTACG | ACTTCCGAGA | TGTTCTGGGC | ACGTGAGTCC  | AGGGCAGGAT  |
| 3101 | TGGGTGCTGG | ATGGCTGAGG | GAGGCTGAGT | CCAGGGTGGG  | GCTTCCTCTG  |
| 3151 | GTCAATTAAT | GCTTCCTGTT | TCCCACAGCC | CAGGCCCTGT  | GGCAGCACTA  |
| 3201 | TCTAGGGCCT | AAACTGTCCC | CAGCTTTTCA | CTTCTGGATG  | ACAGTGGGTG  |
| 3251 | GGACACGGGC | TGCTCTCCCA | ATAGCCCTGG | GTTCTTGAAG  | AGAAAGAAGT  |
| 3301 | CGAGAGAATG | AAGGTGCCAG | TCAGTCCATT | TAACTTGCTG  | CCAAGAGCTA  |
| 3351 | AGTGTCTAG  | CCTAGGTTTG | GGAAGTGGG  | CTGGAGATGG  | CTCTGTTCTT  |
| 3401 | GGTGCTGGGA | ATGCAGAAAT | AACTCAAACC | TGGTCTCTGC  | CCTTCAAGTT  |
| 3451 | GATCCCAGAC | ATGTGCAAGA | GACAGACCTA | CAGAAAATGA  | CAACAGGGTG  |
| 3501 | TGTGCTGTGC | TCCAATTAAG | GTTGGGATTG | AGGGCTTTGT  | GGAGCCCAGA  |
| 3551 | GAGAGCTGTG | CCTTCTGCCT | GGGGGAAAAC | TTCTTGAGAG  | ATGGGGCATT  |
| 3601 | AGAGCTGGGG | ACTGAAGGAT | GGGTAGGTGT | GCACTTGTCA  | GAGAGGAAGA  |
| 3651 | AGGACATTCC | AGGCAGAAGG | AATAGCATAA | ACAAAGGCTT  | AGAGGCATGG  |
| 3701 | TTCTATGTGG | AGAGAGGTAG | AGTGTGATGG | AGCTTAAAT   | CACAGGCTGG  |
| 3751 | GGGGAGAGTG | GAAAAAGGGG | CTGGAGATGA | AAGTGGGACA  | GTTTGTGTAG  |
| 3801 | GGTTTTGGAA | GCCAGGCCAG | GGAGGCTGGA | TATTGTCCCA  | TAGGCCACCG  |
| 3851 | GGAGACACTT | AAGACTTTTT | GGCAGGTGTG | CAATTCAGGA  | TAGTCACTCT  |
| 3901 | GGCCACAGCT | TGGAGGGTAA | ATTGGAGAGG | GACAAGACTG  | GAAACCACTG  |
| 3951 | ATGAGGTTAC | TACAGTAACT | AATTATCCCT | GAGGATTGAA  | ATTTCAACCAC |
| 4001 | GAGAGATGCT | TTTCTTTGAC | TTATGACTTC | TTATTCTCCC  | AGAGAAAGCA  |
| 4051 | AACAGATGTG | GAAAGAATAC | CCTAGCAAAT | CCTCTTTAAT  | CAGTTAACTT  |
| 4101 | TAGTTAAATG | AGTTTATTTG | TTCCTTTTTA | AGAACCTGTT  | CTAAACACT   |
| 4151 | GCTTCTTAAA | GTTCAATGAG | CATACAAATC | ACCTGAGGAT  | TTTGTTAAAC  |
| 4201 | TGCAGATTGA | TTTAGTAAAT | CTGGGGCAGG | GCCTAAAGTT  | TTGCATTTCT  |
| 4251 | TTTTTTTTTC | TTTTTTTTGA | CCCAGGATCC | AAAGCAGTAG  | AGATTTTGCA  |
| 4301 | TTTCTAAAAA | AGTTCCCGGG | TGATGCTGAT | GGTTCTTTAA  | GGTTCTAAAG  |
| 4351 | GGTGTTAAAT | TAGCCATGAC | TCGAATTAGC | AGAAAAAGGG  | ATGAACCAAC  |
| 4401 | TGTACACATA | ATCCAAAAGC | CCAGGGGTAG | ACCTCAGGCA  | TGGCTGGATC  |
| 4451 | CAGAGGGCCA | CATAATGTTA | TCAGGAAATA | TATTTGGCCA  | TTTCTCAGGT  |
| 4501 | TGGACTTCCT | TTGTGTTAAT | TTCATTCCCA | AGCAGGCTCT  | CCCCAGGTGG  |
| 4551 | TGGCAAAGAT | GATCGCCATT | AGCTCCAGGC | TTACATCCTA  | CCAGCTCAAC  |
| 4601 | AGGAGACTCA | TTCTCAAAGT | GCTAGTAAGC | TGGCTTGCA   | CACATGACCA  |
| 4651 | ATTACTGTGG | CCAGGGGAGA | GACTACTTTG | ACTGGCCAGG  | CCTGGGTCAT  |
| 4701 | GTGACCATCT | CTGGAGCCAG | GGGATGGATG | AGTGACTAGG  | GGAGGGTCAT  |
| 4751 | CCACGTCCTT | GGTCCAGCAG | TGGTCACAGA | ACCCATAGGG  | AATGGAGGAG  |
| 4801 | AGGCTGGAGG | GAAGCTGGGG | TTCCAGTTCT | TCACCTTGTTG | AATCCCCTCT  |
| 4851 | CCCAGTAGGG | GGGCCTTCTC | GGAGGTGATC | CTGGCAGAAG  | ATAAGAGGAC  |
| 4901 | GCAGAAGCTG | GTGGCCATCA | AATGCATTGC | CAAGGAGGCC  | CTGGAGGGCA  |
| 4951 | AGGAAGGCAG | CATGGAGAAT | GAGATTGCTG | TCCTGCACAA  | GTGCGTGGGC  |
| 5001 | CACAGCCTTT | CCCTGCCCCA | AGCTGACCCT | GCCTTGGCCC  | TCCCATCCTC  |
| 5051 | CTCCTTTCTT | GCTTTGGACA | AATCATTTAA | ACTCTCTAAG  | CCTTAAATTG  |
| 5101 | CCCCTTTATA | AAATGGGGAT | CACAATTTCC | ACTTGGCAGG  | GTTGTGGGGA  |
| 5151 | ACATCAGAAG | TCCTTTATTT | CAAGTGCCTG | GCCTAACATG  | ACAGATGTGA  |
| 5201 | TGGAGGTGCC | AGTGCTTAGT | CACAGGGGTT | TAACTGTTCA  | ATCAGGTGTA  |
| 5251 | AAGATCCATC | CTGAACATGG | CTTGGACCCA | CATATCTCAG  | TTGGTGTGTG  |
| 5301 | CTCTGGACCT | ACCTCAAGTT | CCCCTCACAT | ATTAACCA    | CTCAGCAAGT  |
| 5351 | TTAAAAATGA | CTGTCTGCTG | ACCCCCAGAC | TAAATCCACA  | ACCAACTGGT  |

FIGURE 3B





|       |             |             |            |            |            |
|-------|-------------|-------------|------------|------------|------------|
| 8101  | CACGCCTGTA  | ATCCCAGCAC  | TTTAGGAGGC | AGAGACGGGT | GGATCACCTG |
| 8151  | AGGTCAGTTC  | GAGACCAGCT  | TGGCAAAACC | CTGTCTCTAC | CAAAAATACA |
| 8201  | AAAATTAGCC  | CGGCGTGGTG  | GCAGATGCCT | GTAATCCCAG | CTACTAGGGA |
| 8251  | GGCTGAGGCA  | GGCGAATCAC  | TTGGACCCGG | GAGGCAGAGG | TTGCAGCAAG |
| 8301  | CTGAGATTGT  | GCCATTGCAC  | GCCAGCCTGG | GCGACAGAGC | GAGCCTGGAC |
| 8351  | GACAGAGCGA  | GACTCCATGT  | CAAAAAAAT  | AAAATAAAAA | CAAAAAATCC |
| 8401  | TATTCCCCTT  | CTGTAGAAAA  | CTTGGATGGG | ACAGCAAAAC | ATAAAGAAAA |
| 8451  | AAGCCAGAAA  | TCCCCGAAAT  | CCTACTCCTC | GGAAATAGCG | ACGGGGCTCA |
| 8501  | CATTTAGCAG  | TACATCTCAA  | TCCGTTCTAG | GAGAAGGGCA | CTTGGGGTGT |
| 8551  | GACATGCCTG  | GTTTTGAATT  | CTGGCTCTGC | TACTGCCTAA | CTGTGGGTTC |
| 8601  | TTGGGTGAGT  | CACTTTGCCT  | CCAAAGGCAT | CAGTTTCCTC | ATCTGTTAGG |
| 8651  | TGAGATTATA  | CAGACTGGCC  | TAGCAGGGAA | GCAGTGAGGA | TGGCATTAAA |
| 8701  | TCAAGCACTA  | ATCCAGGGTC  | TGGCATAAAA | TAGGCATTCA | AACATTCTTT |
| 8751  | TAGGGCTTTA  | CAGTGCACAC  | CTGAGGTTTA | GAGACAGTTC | CCCCCACAC  |
| 8801  | CCTCTTGAGC  | CTTGTCTTTC  | CTGGAATTTT | TGGCCTTCTT | GAGAGCTTCC |
| 8851  | TTGATTTTCT  | TATGACAGCC  | ATGAAGCCAC | AGTGGCTTTT | GGGGATCCAT |
| 8901  | TATTTCTCAG  | AAGGTGCTTG  | GAGCGGCAGA | AGGTTCTACC | AGCCTCTAAC |
| 8951  | CATCTCTGAT  | TGCCCCCTTCT | CTTCCCTCCT | GCCCTTCAAG | CCAGAGAATC |
| 9001  | TGCTGTACTA  | CAGCCTGGAT  | GAAGACTCCA | AAATCATGAT | CTCCGACTTT |
| 9051  | GGCCTCTCCA  | AGATGGAGGA  | CCCGGGCAGT | GTGCTCTCCA | CCGCCTGTGG |
| 9101  | AACTCCGGGA  | TACGTGGGTG  | CGGAGGGCCC | TGGGCTGGGG | CTGTGATGGT |
| 9151  | GGGGGGAACC  | AGGAGTTGAA  | GGGCAGAGAT | TTGTCACCAC | CACGTCCTCT |
| 9201  | TCCCTCCACA  | GCCCCCTGAAG | TCCTGGCCCA | GAAGCCCTAC | AGCAAGGCTG |
| 9251  | TGGATTGCTG  | GTCCATAGGT  | GTCATCGCCT | ACATCTTGTA | AGTGGGGCTT |
| 9301  | GGCCATGGTA  | GGCTGTGGCT  | CCAGAGTTGT | CCTCTCGCCT | ACTTTCCTCT |
| 9351  | CTTCCTTCTT  | CTGCTCTCCC  | TCTGCCCTCC | CTTCCTTCCC | TCCCTCCCTT |
| 9401  | CCTTCCACCA  | ATCAATTACC  | AGTATTACTT | CATTCAATAG | ATACTATGTT |
| 9451  | TCAAGCACTG  | TGCCAAGCAA  | GCACTGGGGT | AAATTTAGCA | CAGCACAAAC |
| 9501  | CAGACAAAGT  | GCCTGCCCTC  | AGGGAGCTGA | CTTTCTTTCT | AGTAGGGAAG |
| 9551  | ACAGACAATC  | AACAAGTAAA  | TAAATCTACA | AACTGACGTC | AGGTGATAAA |
| 9601  | AATAAAATACT | GTGGAGAAAA  | ACCAAGCAGG | AATAGGGAGA | CGGGGTGATG |
| 9651  | CCATTTTCAGT | AGGGAGGTCA  | GGGAAGGGCT | CGCTGTGGAG | GTGATGACCG |
| 9701  | AGTGGTGAGG  | GAGCCAGACA  | TTGGAGGTGT | GGGGAAAGAG | TGGCATAGGC |
| 9751  | AGAAGCAATG  | GCAAGTGCAA  | AGGCCCTGAG | GAGGGCAAGA | TGGCGGCACA |
| 9801  | TACAAGGAAC  | AGAAAGGATA  | ATGTAGCTAG | AACAGGAGTG | AGCAGGCAGG |
| 9851  | GCTGGTAGAG  | TTTATAAAGG  | GGGAACCTCT | TCCATGGCTC | CTGCCTGACC |
| 9901  | CCTGAGACTG  | CCCCAGTGCT  | CCACCCCGGA | GCCAACGGCA | CCCGAAAGTG |
| 9951  | GAAATGAGGA  | TGAGTTTCTC  | CCTGCCCAGG | CTCTGCGGTT | ACCCTCCCTT |
| 10001 | CTATGACGAG  | AATGATGCCA  | AACTCTTTGA | ACAGATTTTG | AAGGCCGAGT |
| 10051 | ACGAGTTTGA  | CTCTCCTTAC  | TGGGACGACA | TCTCTGACTC | TGGTATTTGG |
| 10101 | GGCTTTGCTT  | TTTTCCCCTG  | GGCCCTGCCT | CTGGTTCTCT | CCTCACCTGC |
| 10151 | TTTGGGGGCG  | GTCTCCCTCC  | TGCCTTCCTT | CTGTGCGATT | TTCCAGCACC |
| 10201 | ACACAAAGAG  | CTGTCTTCGA  | GACCAGACAC | CCTACCCCTT | CTTCCTTCTG |
| 10251 | CTTGGGTACT  | TCCTTCTGCT  | TGGCTCCCAG | AGTGAGAAAC | TAGGCATTCA |
| 10301 | TTTGTTCAAT  | CTTCAAACAT  | AGTCTATTTG | AAAATACCTC | TCCCCTATTG |
| 10351 | ACACCCTAAT  | GTCTAAACAA  | CCACCATAAA | CATTTTCATC | CTCCTTTTGT |
| 10401 | GCCCCCTATT  | AAGAAGCAAA  | CCTGTGAAGC | TACTATCGTT | TATCATCAGT |
| 10451 | GTGAATGCAC  | TGAGATTAGT  | CAAGAACAAC | TTTTTTTTTT | TTTTCTTTCT |
| 10501 | TTTTTGAGAC  | GCAGTCTCGC  | TCTGTTGCCC | AGGCTGGAGT | GCAGTGGCAC |
| 10551 | AATCTCGGCT  | CACTGCAACC  | TCTGTCTCCC | GGGTTCGAGC | AATTCTCTGC |
| 10601 | CTCAGCCTCC  | CAAGTAGCTG  | GGATTACAGG | CGCCACCAC  | CATGCCCGGC |
| 10651 | TAATTTTTTT  | TGTATTTTFA  | GTAGAAACAA | GGTTTCACCA | TCTTGGCCAG |
| 10701 | GCTGGTCTTG  | AACTCCTGAC  | CTCGTGATCC | ACCTGCATTG | GCCTCCCAAA |
| 10751 | GTGCTGGGAT  | TACAGACATG  | AGCCACTGTG | CCCGGCCATA | TGTTTTTCTT |

FIGURE 3D

|       |            |            |            |            |             |
|-------|------------|------------|------------|------------|-------------|
| 10801 | AAGAGAGAAA | GGAAAGAGCT | GGAAGGCACG | GGGTGGGAGG | GCCTGAAGAA  |
| 10851 | GAGCATAGGT | TGGGTGGGGT | GGGGCATGGA | CTGATTTGGC | CTCTTTGTCT  |
| 10901 | TGATGCCAGG | CCAGACCTGA | GGGAGTGGGT | ATGCTCTTGG | GGAGTACACA  |
| 10951 | GGCAGTACCA | TGCTGTCATT | ATCTTTGCTT | TTGTCTTGGG | GGTTTAGCCA  |
| 11001 | AAGATTTTAT | CCGGCACTTG | ATGGAGAAGG | ACCCAGAGAA | AAGATTACCC  |
| 11051 | TGTGAGCAGG | CCTTGCAGCA | CCCATGGTGA | GAATTCACAC | AACCTGTGAG  |
| 11101 | CTGGGGCGGG | ATTTGGGGCC | CTCAGGTCTG | CTTCTGCCCT | CATAGGCAAC  |
| 11151 | CCACCACATA | ACCCCATCCT | AGGATTGCAG | GAGATACAGC | TCTAGATAAG  |
| 11201 | AATATCCACC | AGTCGGTGAG | TGAGCAGATC | AAGAAGAACT | TTGCCAAGAG  |
| 11251 | CAAGTGGAAG | GTGAGTCCAT | ATCCCTAGTT | CTGGTCCCAG | CCTCCCCCAGG |
| 11301 | ACTCCTCCCC | ATCCCTACCC | AGGCTCAGCT | TGCACAGCAC | CTGGCATCAC  |
| 11351 | ACTGGGCACA | CAGTAACTGC | TTAGGGATCC | TTACTGAAGG | ACTTCATTCA  |
| 11401 | TTCACTCTTT | CATTCAACAA | ACACTCCCAA | CACCTTCTCT | ATTCCAGAGA  |
| 11451 | GGGTCCCTCA | CCTCCAAGTC | TAGAGGAAGA | AGTCTGTAAT | TCTTCAGGAG  |
| 11501 | GCATCTGATC | CAGCCTATGG | GGTCCGAGAA | AGGTCATAAA | AGTGGTGATG  |
| 11551 | ACCTGACAGA | GCTGTCAGTT | AAGTAGGAAT | TAGTGAGGCA | TAGCGGAATA  |
| 11601 | ATGTCTATAG | CCATTCCGGG | AAGTGCAAGT | GCTAAGCCTG | GCCAGACTGG  |
| 11651 | AGGGGCTGAG | GGGACTGAGA | GGCAGGAGCC | CAATTTAGAG | AAGCAGGTAA  |
| 11701 | GGGGCCAGGC | CTCTTAGGGC | CTCATATGCC | ACAGAGGAGC | ACCAACTTGA  |
| 11751 | TCCTGAGGGC | ACTGAGGAGC | CCCAGAAGAA | TCTTAGGCAA | GTATTTGCTG  |
| 11801 | CATAGAAAGG | GCTCTCAGGG | CCAGGCATGG | TGGCTCACGC | CTGTAATCCC  |
| 11851 | AGCACTTTGG | GAGGCCGAGG | TGGTTGGATC | ACCTGAGGTC | AGGAGTTCAA  |
| 11901 | GACCATCCTG | GCCAACATGG | CAAAACCCTG | TCTCTACTAA | AAATAAAAGA  |
| 11951 | ATTAGCCACA | CATGGTGGTG | CGTGCCTGTA | ATCCCAGCTA | CTTGGGAAGC  |
| 12001 | TGAGGCAGGA | GAATTACTTG | AACCTCGGAG | ATGGGGGTTG | CAGTGAGCTG  |
| 12051 | AGATCGCGCC | ACTGCACTCC | AGCCTGGGCA | ACAAAGTGAG | ACTCCACCTC  |
| 12101 | AAAAAAAAAA | GAAAGAGCTC | TCAGGATGCA | GAGAATGGCA | TGGAGTAAAG  |
| 12151 | ACTGGGTGAC | GCATTAGGAG | GCTGTGGCAG | AGATACAGGC | AGGAGATGGT  |
| 12201 | AAGGGTTTGG | AACCACAGTA | GCAGCAACAG | GGGGCAGAGA | ACAGTGGTTG  |
| 12251 | ATCCAGGAGT | CATTTAGGAG | GTGAAACTGA | CAAGACATGA | CGATGCAATG  |
| 12301 | GATGTTGGGG | GAAAGAGATG | TCAAGGGCTG | GCCCAAGACT | GTGGCTGGGA  |
| 12351 | ACAGAAATGA | TGGTGGTGGT | ACCATGACTG | AGATGGTTAT | CACAGGGACA  |
| 12401 | GAAACATGTT | TTGGGGGGAT | GGTTTTAGTT | TTAGACATGG | TGAATTTGAG  |
| 12451 | GGGTGTGTGG | GACACCTAGG | TGGAGATATT | GAATAGAGAC | ACACCTGAGC  |
| 12501 | AAGTTACTTC | AGCTTTCTGT | GCCTCAGTTT | CCTCCTTTGA | AAATGATAAT  |
| 12551 | AGTACCTACC | TCAAAGACTT | TCATGAAGAT | TAAATGAATT | ACTACGTAAG  |
| 12601 | GTGCTTAGAA | CAGTGCCTGA | CATACAGTGC | TATAGTGTTT | GCTATTACAT  |
| 12651 | ATTAATATGA | ATTATAGTTA | TGTTTCTATT | TATATATATA | GATACACATA  |
| 12701 | CATCTAACAT | ATGTGCGTGT | GTGTGTGTAA | ATATATAATA | AAGCCTTGTA  |
| 12751 | GAGGTTTTTG | GGGGGCTTTA | GGGGAATTAA | TAAAATAACT | CCTGAATGAA  |
| 12801 | AATAACAGAA | CAATTGCAAG | AATCCCCTG  | CGCCCCTGCC | CCATGACTTG  |
| 12851 | ACTCTCTCAA | AAGTCCTTTC | TCCCCTCTCC | CTTCAATGCC | TTCAATGCCA  |
| 12901 | GCAAGCCTTC | AATGCCACGG | CTGTGGTGCG | GCACATGAGG | AAACTGCAGC  |
| 12951 | TGGGCACCAG | CCAGGAGGGG | CAGGGGCAGA | CGGCGAGCCA | TGGGGAGCTG  |
| 13001 | CTGACACCAG | TGGCTGGGGG | TGAGGAGCGG | GCTCTGCAGA | AGGGCATGGG  |
| 13051 | TGGTCCACAA | AGGTGCACCC | GGGCTGGAGT | GGAGGGCCTG | CCCCTGCGGC  |
| 13101 | CACCTCTGTT | CTGTCTTCCC | ATGCAGGGCC | GGCAGCTGGC | TGTTGCTGTC  |
| 13151 | GAGACTGCTG | CGTGGAGCCG | GGCAGAGAAC | TGTCCCCCAC | ACTGCCCCAC  |
| 13201 | CAGCTCTAGG | GCCCTGGACC | TCGGGTCATG | ATCCTCTGCG | TGGGAGGGCT  |
| 13251 | TGGGGGCAGC | CTGCTCCCCT | TCCCTCCCTG | AACCGGGAGT | TTCTCTGCCC  |
| 13301 | TGTCCCCTCC | TCACCTGCTT | CCCTACCACT | CCTCACTGCA | TTTTCCATAC  |
| 13351 | AAATGTTTCT | ATTTTATTGT | TCCTTCTTGT | AATAAAGGGA | AGATAAAACC  |
| 13401 | ATCCTTAGCG | CTGTCTCCCT | CAATATCCCC | CACCCCATCT | TGTTGTGCAA  |
| 13451 | ACTGACTGCT | TGATTTGGGG | GTGCCTGGCC | TTTGAGGTAG | TCACAGGGAG  |

FIGURE 3E

|       |            |            |            |            |             |
|-------|------------|------------|------------|------------|-------------|
| 13501 | CCCCCTCCCC | AACATGAGAC | TGGGTGGGGA | TGGGGAGAGA | GAAGTGGGGA  |
| 13551 | ATGGAGGGGA | AGGTGCTTGG | GGAATTTCTT | TGTCCAGGGT | GCCCCATCTA  |
| 13601 | GCCTTCCGGC | CCTTTGGAAC | CCTTTCTGCG | CTTTGCTGGT | GGCTCCTGAG  |
| 13651 | CATGGCGGGA | TTGGCGCAGG | TCGGCACTGA | ACAGCACCTG | TAGGAGGGTG  |
| 13701 | GAGTCTGTGT | GGGGAGGAGG | GTACACTGGG | GTCAGGGCTG | GTGAGACTAG  |
| 13751 | TGACAGTGTT | GGGAGGTGGA | AGAGTCCTTG | GGGAACAGGG | CCGAAGGCCAA |
| 13801 | TGAGAATCCA | CTGGGGTTGG | GACAGGGGTG | GCTGGAGAGT | CCTTTAGGGC  |
| 13851 | CACCTGGGGC | GGTGGTGGAA | GAGTCCACTG | GGTCTGGGCT | GGAGGAGAGG  |
| 13901 | AAACCTAGGG | AGGACACCTA | GGTACACTCA | CCGCTTGGGC | CCAGCCAGCA  |
| 13951 | TAAGGTCCCC | ACAGGCTCCG | GAAAAAGTTT | CCTAAATCAG | AAGTGATGAG  |
| 14001 | ACTAAGTTAT | CTGACCCCTT | CTGTGACCCA | TCAACAGAAG | TAGGGTCTGA  |
| 14051 | GGGAGAGGTG | ACTAAGAGAG | AGAGAAGTTT | CTACCATCCC | AGCCCACTGC  |
| 14101 | CAGCCCCTGC | AGCCCACTTT | CCTCACCCAG | TTCTTGTTG  | GTCTGGGGGC  |
| 14151 | TCGGTCCCTT | CGCCTGGGAC | GTGGTAGGGT | GCCAGCTGTA | GTCACGTTGG  |
| 14201 | GCAATGTGCC | ACATATGGAC | ATCCACGGGC | ACAGCCTGGG | GCTTGTCTAG  |
| 14251 | GGCCATCAGG | CAGATGCAGT | CAGCCACCTT | TGACAGACAC | AGAATGAGCC  |
| 14301 | CTTGTGGAAG | AAGGGCAGCA | TGTGGCCAGC | ATCTTGCTTA | TAGCCCCAAA  |
| 14351 | GCCGGCTGCT | TTCTCCTTCA | CTCTGGGGTT | ACTGTTGTTT | TATATTCTCA  |
| 14401 | ATCAACAGAT | ACTATCTATG | AATACACTTT | TTTTTTGTTT | GTTTTTGAGA  |
| 14451 | TGGAGTCTCG | CTCTGTTGCC | TAGGCTGGAG | TGCACTGGTG | CAATCCTGGC  |
| 14501 | TCTCCCAGGT | TCAAGCAATT | CTCCTACCTC | AGCCTCCCAA | GTAGCTGGGA  |
| 14551 | TTACAGGCAT | GTGCCACCAC | GTGTGGCTAA | TTTTTGTTGT | TTTAGTAGAG  |
| 14601 | ATGGGGTTTC | ACCATGTTGG | CCAGCCTGGT | CTCGAACTCC | TGACCTCAAG  |
| 14651 | TGATCTGTCC | ACCTTGGCCT | CCCAAAGTGC | TGGGATTACA | GGCGTGAGCC  |
| 14701 | ACCATGCGCG | GCCTATGAAT | ACACTGAAAT | TGCTGTAATA | AGAGGTGCTA  |
| 14751 | CTAGCTGAAC | ACCTATGTGG | GCCAGGTTAT | CATAACCTGG | GAAGAAGGTA  |
| 14801 | TTACCACACC | CACCTTACAG | ACAAGAAAAC | TGAGGCTTTG | AAAGGTGAAG  |
| 14851 | TGACCTGGCC | AAAGTCACAT | GGCTGAGAAT | AGGCAGAACC | AAGATTTAAT  |
| 14901 | GTTAGGCTGT | AGTCCAAAGC | CCATCAAAAA | AAAATCTTTA | AGCAAAAATT  |
| 14951 | CATTTTTTAA | ACTACAGAGA | AGTATAAAGA | AAAAAAAAGG | CTGGGTGCAG  |
| 15001 | TGGCTCACGC | CTGTAATCCC | AGCACTTTGG | GAGGCTGAGG | CAGGTGGATC  |
| 15051 | TCGAGGTCAG | ATTGAGACCA | TCCTGGCCCA | ACATGGTGAA | ACCCCATCTC  |
| 15101 | TACTAAAAAT | ACAAAAATTA | GCTGGGTGTG | GTGGCGCATG | CCTGTAATCC  |
| 15151 | CAGCTAATCT | GGAGGCTGAG | GCAGGAGAAT | AGCTTGAAGC | CGGGAGGCGG  |
| 15201 | AGGTTGCAGT | GAGCCGAGAT | TGCACCACTG | CACTCCAGCC | TGGCAACAAA  |
| 15251 | GCAAGACTCC | ACCTCAAAAA | AAAAAAAAAA | AAGACAAATG | CCTAATTTCC  |
| 15301 | AGTCATCTTA | TTGCCAGTTA | ACCCTATTGA | CATCAAGCAA | AAAGTTTTGT  |
| 15351 | CAGTACATGT | CATTTTACGA | AAGGAACAAA | ATGTGGCCGG | GAGCAGTGCC  |

(SEQ ID NO:3)

# FEATURES:

## Genewise results:

Start: 3000  
Exon: 3000-3082  
Exon: 7535-7609  
Exon: 7696-7834  
Exon: 8991-9117  
Exon: 9212-9287  
Exon: 9980-10092  
Exon: 10998-11076  
Exon: 11173-11260  
Exon: 12902-13019  
Exon: 13127-13206  
Stop: 13207

FIGURE 3F

#### Sim4 results:

Exon: 3000-3082, (Transcript Position: 1-83)  
Exon: 7535-7609, (Transcript Position: 84-158)  
Exon: 7696-7834, (Transcript Position: 159-297)  
Exon: 8991-9117, (Transcript Position: 298-424)  
Exon: 9212-9287, (Transcript Position: 425-500)  
Exon: 9980-10092, (Transcript Position: 501-613)  
Exon: 10998-11076, (Transcript Position: 614-692)  
Exon: 11173-11260, (Transcript Position: 693-780)  
Exon: 12902-13019, (Transcript Position: 781-898)  
Exon: 13127-13209, (Transcript Position: 899-981)

CHROMOSOME MAP POSITION:  
chromosome 3

#### ALLELIC VARIANTS (SNPs):

| DNA<br>Position | Major | Minor | Domain |
|-----------------|-------|-------|--------|
| 1892            | A     | G     | Intron |
| 3351            | G     | A     | Intron |
| 8636            | T     | A     | Intron |
| 8805            | T     | C     | Intron |
| 9802            | T     | A     | Intron |
| 9833            | C     | G     | Intron |
| 11352           | C     | T     | Intron |
| 13319           | T     | C A G | Intron |
| 13659           | C     | G     | Intron |
| 14292           | G     | C     | Intron |

Context:

DNA

Position

1892 GGTGTTTCGGGTGCAGGGCTGTCTGGGGCACTGTGTGGTGTGGACATGTGTGCTGATGTCT  
GGGAGTACATGTATGATCAGGTGTACGGGATGTGGATAACAAGGCGTACTGGATCTGGGA  
GGCAGGTGTTTGAGTTCAGGGCTGTGGAGGGGGCTTGGTGTGGCATGTCTGCTACAGGGA  
TGTGTGTGGATCTGTGAGGGTTGTATTTGGTAGGCCTCCATGTGGGTTTCAGACTCTGCC  
TCTAGAGCTTACACTCGAGTCTCCTTTCTAGAAGATTCTGCCCCTGGATGGGTGGGAG  
[A,G]

GTCCCCTGGGAAAAAGGTCCTGTTCCAGGAGTGGAATCTCACACCAGAGGCCCTAGTCAG  
GGCACCTTCTCCTCATTTCTCCCTTAGAGAAAAAGAGAGAAGGAAAGTGCTCTCCCTGAGG  
TCACAAAGCATGCTGGGCTCTGTTTTGGCCTCATCTGTGGATGGGTGGGAGGCTGTGTT  
CTCTGAATGGGGCCCATTTCTGGCTTCATATTGGAAGTACCAGCCAAGGCCATTCGATGGC  
CTTTGCCCTCAGCAAGCTTAGCTGGGGGGCCCCAGGCCAGGTGTCATTAGGGCCTCTGGAG

3351

GACATCTACGACTTCCGAGATGTTCTGGGCACGTGAGTCCAGGGCAGGATTGGGTGCTGG  
ATGGCTGAGGGAGGCTGAGTCCAGGGTGGGGCTTCTCTGGTCAATTAATGCTTCTGTT  
TCCCACAGCCCAGGCCCTGTGGCAGCACTATCTAGGGCCTAAACTGTCCCAGCTTTTCA  
CTTCTGGATGACAGTGGGTGGGACACGGGCTGCTCTCCCAATAGCCCTGGGTTCTTGAAG  
AGAAAGAAGTCGAGAGAATGAAGGTGCCAGTCAGTCCATTAACTTGCTGCCAAGAGCTA  
[G,A]

GTGTTCTAGCCTAGGTTTGGGAACTGAGGCTGGAGATGGCTCTGTTCTTGGTGCTGGGAA

FIGURE 3G

TGCAGAAATAACTCAAACCTGGTCTCTGCCCTTCAAGTTGATCCCAGACATGTGCAAGAG  
ACAGACCTACAGAAAATGACAACAGGGTGTGTGCTGTGCTCCAATTAAGGTTGGGATTGA  
GGGCTTTGTGGAGCCCAGAGAGAGCTGTGCCTTCTGCCTGGGGGAAAACCTCCTGGAGAA  
TGGGGCATTAGAGCTGGGGACTGAAGGATGGGTAGGTGTGCACTTGTGAGAGAGGAAGAA

8636 AGAGCGAGCCTGGACGACAGAGCGAGACTCCATGTCAAAAAAATAAAAACAAAA  
AATCCTATTCCCCTTCTGTAGAAAACCTGGATGGGACAGCAAAACATAAAGAAAAAGCC  
AGAAATCCCCGAAATCCTACTCCTCGGAAATAGCGACGGGGCTCACATTTAGCAGTACAT  
CTCAATCCGTTCTAGGAGAAGGGCACTTGGGGTGTGACATGCCTGGTTTTGAATCTGGC  
TCTGCTACTGCCTAACTGTGGGTCTTGGGTGAGTCACTTGCCTCCAAAGGCATCAGTT  
[T, A]  
CCTCATCTGTTAGGTGAGATTATACAGACTGGCCTAGCAGGGAAGCAGTGAGGATGGCAT  
TAAATCAAGCACTAATCCAGGGTCTGGCATAAAATAGGCATTCAAACATTCCTTTAGGGC  
TTTACAGTGACACCTGAGGTTTAGAGACAGTTCCCCCCCACACCCTCTTGAGCCTTGTG  
CTTCTGGAATTTTTGGCCTTCTTGAGAGCTTCTTGATTTTCTTATGACAGCCATGAAG  
CCACAGTGGCTTTTGGGGATCCATTATTTCTCAGAAGGTGCTTGGAGCGGCAGAAGGTTG

8805 TAGCAGTACATCTCAATCCGTTCTAGGAGAAGGGCACTTGGGGTGTGACATGCCTGGTTT  
TGAATCTGGCTCTGCTACTGCCTAACTGTGGGTCTTGGGTGAGTCACTTGCCTCAA  
AGGCATCAGTTTCTCATCTGTTAGGTGAGATTATACAGACTGGCCTAGCAGGGAAGCAG  
TGAGGATGGCATTAAATCAAGCACTAATCCAGGGTCTGGCATAAAATAGGCATTCAAACA  
TTCTTTAGGGCTTTACAGTGACACCTGAGGTTTAGAGACAGTTCCCCCCCACACCCTC  
[T, C]  
TGAGCCTTGTCTTCTGGAATTTTTGGCCTTCTTGAGAGCTTCTTGATTTTCTTATGA  
CAGCCATGAAGCCACAGTGGCTTTTGGGGATCCATTATTTCTCAGAAGGTGCTTGGAGCG  
GCAGAAGGTTCTACCAGCCTTAACCATCTCTGATTGCCCTTCTCTTCCCTCCTGCCCT  
TCAAGCCAGAGAATCTGCTGTACTACAGCCTGGATGAAGACTCCAAAATCATGATCTCCG  
ACTTTGGCCTCTCCAAGATGGAGGACCCGGGCAGTGTGCTCTCCACCGCCTGTGGAATC

9802 AGACAAAGTGCCTGCCCTCAGGGAGCTGACTTTCTTTCTAGTAGGGAAGACAGACAATCA  
ACAAGTAAATAAATCTACAACTGACGTGAGGTGATAAAAAATAAATACTGTGGAGAAAA  
CCAAGCAGGAATAGGGAGACGGGTGATGCCATTTAGTAGGGAGGTGAGGGAAGGGCTC  
GCTGTGGAGGTGATGACCGAGTGGTGAGGGAGCCAGACATTGGAGGTGTGGGGAAAGAGT  
GGCATAGGCAGAAGCAATGGCAAGTCAAAGGCCCTGAGGAGGGCAAGATGGCGGCACAT  
[T, A]  
CAAGGAACAGAAAGGATAATGTAGCTAGAACAGGAGTGAGCAGGCAGGGCTGGTAGAGTT  
TATAAAGGGGGAACCTCTTCCATGGCTCCTGCCTGACCCCTGAGACTGCCCCAGTGCTCC  
ACCCCGGAGCCAACGGCACCCGAAAGTGGAAATGAGGATGAGTTTCTCCCTGCCAGGCT  
CTGCGGTTACCCTCCCTTCTATGACGAGAATGATGCCAACTCTTTGAACAGATTTTGAA  
GGCCGAGTACGAGTTTGACTCTCCTTACTGGGACGACATCTCTGACTCTGGTATTTGGGG

9833 TTCTTTCTAGTAGGGAAGACAGACAATCAACAAGTAAATAAATCTACAACTGACGTGAG  
GTGATAAAAAATAAATACTGTGGAGAAAAACCAAGCAGGAATAGGGAGACGGGGTATGCC  
ATTTAGTAGGGAGGTGAGGGAAGGGCTCGCTGTGGAGGTGATGACCGAGTGGTGAGGGA  
GCCAGACATTGGAGGTGTGGGGAAAGAGTGGCATAGGCAGAAGCAATGGCAAGTCAAAG  
GCCCTGAGGAGGGCAAGATGGCGGCACATAACAAGGAACAGAAAGGATAATGTAGCTAGAA  
[C, G]  
AGGAGTGAGCAGGCAGGGCTGGTAGAGTTTATAAAGGGGGAACCTCTTCCATGGCTCCTG  
CCTGACCCCTGAGACTGCCCCAGTGCTCCACCCCGGAGCCAACGGCACCCGAAAGTGGAA  
ATGAGGATGAGTTTCTCCCTGCCAGGCTCTGCGGTTACCCTCCCTTCTATGACGAGAAT  
GATGCCAACTCTTTGAACAGATTTTGAAAGCCGAGTACGAGTTTGACTCTCCTTACTGG  
GACGACATCTCTGACTCTGGTATTTGGGGCTTTGCTTTTTTCCCTGGGGCCTGCCTCTG

11352 GTGAGCAGGCCTTGCAGCACCCATGGTGAGAATTCACACAACCTGTGAGCTGGGGCGGGA

FIGURE 3H

TTTGGGGCCCTCAGGTCTGCTTCTGCCCTCATAGGCAACCCACCACATAACCCCATCCTA  
GGATTGCAGGAGATACAGCTCTAGATAAGAATATCCACCAGTCGGTGAGTGAGCAGATCA  
AGAAGAACTTTGCCAAGAGCAAGTGGAAGGTGAGTCCATATCCCTAGTTCTGGTCCCAGC  
CTCCCCAGGACTCCTCCCCATCCCTACCCAGGCTCAGCTTGACAGCACCTGGCATCACA  
[C, T]  
TGGGCACACAGTAACCTGCTTAGGGATCCTTACTGAAGGACTTCATTCACTCTTTCA  
TTCAACAAACACTCCCAACACCTTCTCTATTCCAGAGAGGGTCCCTCACCTCCAAGTCTA  
GAGGAAGAAGTCTGTAATTCTTCAGGAGGCATCTGATCCAGCCTATGGGGTCCGAGAAAG  
GTCATAAAAGTGGTGATGACCTGACAGAGCTGTCACTTAAGTAGGAATTAGTGAGGCATA  
GCGGAATAATGTCTATAGCCATTCCGGAAGTGCAAGTGCTAAGCCTGGCCAGACTGGAG

13319 GGTGAGGAGCGGGCTCTGCAGAAGGGCATGGGTGGTCCACAAAGGTGCACCCGGGCTGGA  
GTGGAGGGCCTGCCCCTGCGGCCACCTCTGTTCTGTCTTCCCATGCAGGGCCGGCAGCTG  
GCTGTTGCTGTGAGACTGCTGCGTGGAGCCGGGCACAGAACTGTCCCCACACTGCCCC  
ACCAGCTCTAGGGCCCTGGACCTCGGGTCTGATCCTCTGCGTGGGAGGGCTTGGGGCA  
GCCTGCTCCCCTTCCCTCCCTGAACCGGGAGTTTCTCTGCCCTGTCCCCTCCTCACCTGC  
[T, C, A, G]  
TCCCTACCACTCCTCACTGCATTTTCCATACAAATGTTTCTATTTTATTGTTCTTCTTG  
TAATAAAGGGAAGATAAAACCATCCTTAGCGCTGTCTCCCTCAATATCCCCACCCCATC  
TTGTTGTGCAAACTGACTGCTTGATTTGGGGGTGCCTGGCCTTTGAGGTAGTCACAGGGA  
GGCCCCTCCCCAACATGAGACTGGGTGGGGATGGGGAGAGAGAAGTGGGGAATGGAGGGG  
AAGGTGCTTGGGGAATTTCTTTGTCCAGGTGCCCATCTAGCCTTCCGGCCCTTTGGAA

13659 CTATTTTATTGTTCTTCTTGTAATAAAGGGAAGATAAAACCATCCTTAGCGCTGTCTCC  
CTCAATATCCCCACCCCATCTTGTTGTGCAAACTGACTGCTTGATTTGGGGGTGCCTGG  
CCTTTGAGGTAGTCACAGGGAGGCCCTCCCCAACATGAGACTGGGTGGGGATGGGGAGA  
GAGAAGTGGGGAATGGAGGGGAAGGTGCTTGGGGAATTTCTTTGTCCAGGTGCCCATC  
TAGCCTTCCGGCCCTTTGGAACCCCTTCTGCGCTTGCTGGTGGCTCCTGAGCATGGCGG  
[C, G]  
ATTGGCGCAGGTGGCACTGAACAGCACCTGTAGGAGGGTGGAGTCTGTGTGGGGAGGAG  
GGTACACTGGGGTCAAGGCTGGTGAGACTAGTGACAGTGTGGGAGGTGGAAGAGTCTT  
GGGGAACAGGGCCGAAGGCAATGAGAATCCACTGGGGTGGGACAGGGGTGGCTGGAGAG  
TCCTTTAGGGCCACCTGGGGCGGTGGTGGAAGAGTCCACTGGGTCTGGGCTGGAGGAGAG  
GAAACCTAGGGAGGACACCTAGGTACACTCACCGCTTGGGCCAGCCAGCATAAGGTCCC

14292 AGTGATGAGACTAAGTTATCTGACCCCTTCTGTGACCCATCAACAGAAGTAGGGTCTGAG  
GGAGAGGTGACTAAGAGAGAGAGAAGTTTCTACCATCCCAGCCCACTGCCAGCCCCTGCA  
GCCACTTTCTCACCCAGTTCTTGTGGTCTGGGGGCTCGGTCCCTTCGCTGGGACG  
TGGTAGGGTGCCAGCTGTAGTCACGTTGGGCAATGTGCCACATATGGACATCCACGGGCA  
CAGCCTGGGGCTTGTCTAGGGCCATCAGGCAGATGCAGTCAGCCACCTTTGACAGACACA  
[G, C]  
AATGAGCCCTTGTGGAAGAAGGGCAGCATGTGGCCAGCATCTTGCTTATAGCCCCAAAGC  
CGGCTGCTTTCTCCTTCACTCTGGGGTACTGTTGTTCTATATTCTCAATCAACAGATAC  
TATCTATGAATACACTTTTTTTTTGTTTGTGTTTGGAGATGGAGTCTCGCTCTGTTGCCTA  
GGCTGGAGTGCACTGGTGCAATCCTGGCTCTCCAGGTTCAAGCAATTCTCCTACCTCAG  
CCTCCAAGTAGCTGGGATTACAGGCATGTGCCACCACGTGTGGCTAATTTTTGTGTTTT

FIGURE 3I